

Sur la production de logiciels libres au Laboratoire d'Informatique Gaspard-Monge (LIGM) : ce que nous avons appris

Teresa Gomez-Diaz

CNRS - Laboratoire d'informatique Gaspard-Monge

Cette œuvre est mise à disposition selon les termes de la licence
Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Journées Mathrice 2021

IMJPRG, Paris, 19-21 octobre 2021





– Un cas d'étude : le LIGM

- LIGM : Laboratoire d'informatique Gaspard-Monge
- Tutelles : CNRS, ENPC, Université Gustave Eiffel, à l'est de Paris
- <http://ligm.u-pem.fr/>
- environ 150 membres, dont 80 chercheurs permanents
- thèmes de recherche : *l'analyse et le traitement des images, la géométrie en informatique, l'apprentissage, les réseaux, l'algorithmique temps-réel, la combinatoire, l'algorithmique, la bio-informatique, la logique informatique et les automates, le traitement automatique des langues, le traitement du signal.*
- Production importante de logiciels, dont ~ **80% FLOSS** :
 - ▶ **finalité** : principalement de la recherche
 - ▶ 2008-13 : 61 logiciels de la recherche, dont 53 FLOSS
 - ▶ 2013-18 : 66 logiciels de la recherche, dont 50 FLOSS
 - ▶ certains très connus :
Unitex (linguistique), OpenMVG (images), contributions à JAVA (JSR 335, JSR 292), MesoNet (Deepfake)...

La mission logiciels de la recherche au LIGM

- 2002 : Ingénieure de recherche CNRS
- 2006 : Mission logiciels (de la recherche) LIGM

*étudier la mise en place de services (suivi de versions, publication, etc.)
associés au développement logiciel au sein du laboratoire, avec
pour objectif de favoriser leur visibilité.*

- 2008-13 :  – Projet PLUME, projet national CNRS

(*)(2019) TGD : Le Projet PLUME et le paysage actuel des logiciels de la recherche dans la science ouverte

- ▶ publication des fiches descriptives de logiciels
- ▶ études sur FLOSS, questions légales : licences et droits d'auteur, procédure de diffusion, politiques de diffusion et FLOSS...

Patrimoine logiciel d'un laboratoire : <https://projet-plume.org/patrimoine-logiciel-laboratoire>

- 2017-18 : Projet PRESOFT, Plans de gestion de logiciels

(2018) TGD, G. Romier : Research Software management Plan Template V3.2, Projet PRESOFT

- depuis 2018 : des travaux sur les logiciels de la recherche, leur évaluation, la Science Ouverte, les infrastructures pour la Science Ouverte, voir :

(2019) TGD, T. Recio : On the evaluation of research software: the CDUR procedure

(2020-21) TGD, T. Recio : Towards an Open Science definition as a political and legal framework: on the sharing...

(2021) TGD, T. Recio : Open comments on the Task Force SIRS report: Scholarly Infrastructures for Research Soft.

(*) Les références complètes sont à la fin de la présentation.

Avec un grand MERCI ! à Mathrice

- 2007 : *Autour de la valorisation de logiciels d'un laboratoire de recherche*
Journées Mathrice, IHP - Paris, 9-11 octobre 2007
<http://igm.univ-mlv.fr/~teresa/logicielsLIGM/documents/Mathrice/talkmathriceOct2007.pdf>
- 2009 : *Autour de la valorisation de logiciels : laboratoire de recherche et projet RELIER*
Journées Mathrice, LAREMA - Angers, 17-19 mars 2009
<http://igm.univ-mlv.fr/~teresa/logicielsLIGM/documents/Mathrice/talkmathriceMars2009.pdf>
- 2012 : *PLUME*
Journées Mathrice - Orléans, 9-11 octobre 2012
<http://igm.univ-mlv.fr/~teresa/logicielsLIGM/documents/Mathrice/talkmathriceOct2012.pdf>
- 2019 : *Projet PRESOFT : plans de gestion de logiciels de la recherche*
Journées Mathrice à l'IHP, Paris, 20 mars 2019
<http://igm.univ-mlv.fr/~teresa/logicielsLIGM/documents/Mathrice/talkmathriceOct2019.pdf>
- 2021 : *Sur la production de logiciels libres au Laboratoire d'Informatique Gaspard-Monge (LIGM)*
Journées Mathrice 2021, IMJPRG, Paris, 19-21 octobre 2021
<http://igm.univ-mlv.fr/~teresa/logicielsLIGM/documents/Mathrice/talkmathriceOct2021.pdf>

Les problèmes détectés en 2006-07 : vision 2021

Sésame, ouvre-toi ! Quels sont les logiciels du LIGM ?

→ produire la liste des logiciels

Définition Ce que je fais est un logiciel LIGM ?

→ besoin de définir “logiciel du laboratoire” ou “log. de la recherche”

Référence Pas de nom, pas d’auteurs, pas de date, pas de version...

→ besoin d’établir une référence (identification)

Diffusion Quelques uns sur pages web perso, sur des forges, avec page web...

→ besoin d’établir une procédure de diffusion

FLOSS Un réel intérêt pour les logiciels libres, mais souvent pas de licence

→ besoin de mieux comprendre : FLOSS, licences

Questions légales Qui décide de la licence ?

→ besoin de mieux comprendre : questions légales, droit d’auteur

Questions de politique Définition, licences, diffusion, décisions...

→ besoin d’établir les politiques du laboratoire sur cette production

Valeur Quelle est la valeur de cette production scientifique ?

→ besoin d’établir des protocoles d’évaluation de la recherche

Questions de nature diverse : juridique, politique, publication, scientifique, bonnes pratiques...

nécessitant des approches différentes, mais similaires pour tout domaine scientifique.

Définition d'un logiciel de la recherche

(2007) TGD : Autour de la valorisation de logiciels développés dans un laboratoire de recherche

(2009) TGD : Guide laboratoire pour recenser ses développements logiciels

(2011, 2015) TGD : Article vs. Logiciel : questions juridiques et de politique scientifique dans la production de logiciels

Un logiciel du laboratoire est un programme utile pour faire avancer la recherche qui a été produit avec la participation d'un membre du laboratoire.
Il arrive souvent que des publications de recherche soient associées.

- finalité : recherche
- un membre du laboratoire participe à l'écriture du logiciel (idem aux publications)
- la production importante sont les articles, les logiciels sont des objets associés

(2019) TGD, T. Recio : On the evaluation of research software: the CDUR procedure

(2021) TGD, T. Recio : Open comments on the Task Force SIRS report: Scholarly Infrastructures for Research Software

Research software (RS) is a well identified set of code that has been written by a well identified research team. It is software that has been built and used to produce a result published or disseminated in some article or scientific contribution.

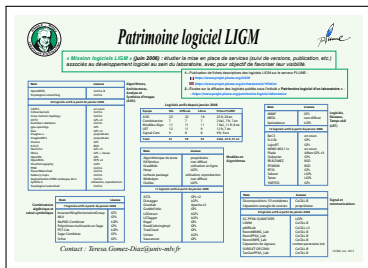
Each RS encloses a set of files containing the source code and the compiled code. It can also include other elements as the documentation, specifications, use cases...

Références, Publications, Liste des logiciels LIGM

– LIGM 2008-2013 (PLUME) : 61 logiciels, dont 53 FLOSS



Metadata, mots clés, classification,
interfaces de recherche... et liens aux articles.



– LIGM 2013-2018 (document interne) : 66 logiciels, dont 50 FLOSS

– Évolutions

- Groupe Science Ouverte de l'Université Gustave Eiffel, voir (2020-21) TGD, T. Recio : Towards an Open Science definition as a political and legal framework: on the sharing...
- (depuis 2011) Image Processing On Line (IPOL) Journal : article + logiciel
P. Monasse, Extraction of the Level Lines of a Bilinear Image, IPOL 2019,
<https://www.ipol.im/pub/art/2019/269/>
- (depuis 2015) SoftwareX Journal, avec le logiciel sur GitHub
B. Perret, et al., Higma : Hierarchical Graph Analysis, SoftwareX, Vol. 10, 2019,
<https://www.sciencedirect.com/science/article/pii/S235271101930247X>

Visibilité et interfaces de recherche

- Comment construire la visibilité des logiciels LIGM ?
- Difficile de trouver un logiciel si nous ignorons : nom, équipe de développement...
- L'information scientifique doit être scientifiquement classifiée.
- Voir (2013) TGD : PLUME et les logiciels de recherche, Séminaire Cocin,
<http://igm.univ-mlv.fr/~teresa/logicielsLIGM/documents/Seminaires/2013marsParisCocin.pdf>

Types d'interfaces de recherche :

- Recherche textuelle simple : Recherche
- Classification thématique (expert thématique)
- Recherche à facettes, par exemple (voir [Wikipédia : Recherche à facettes](#)) :
 - ▶ selon la date de publication (ou modification) des résultats,
 - ▶ au type des résultats (en général, type des documents),
 - ▶ la langue,
 - ▶ présence de mots ou d'expressions clés liés à la recherche,
 - ▶ présence d'entité nommée : personne, un lieu, une organisation...
- Mots clés de la plateforme : classification en sous-thématiques
- Mots clés Laboratoires et Institutions : Logiciels du LIGM, du CNRS...
- Mots clés libres (choisis par les déposants de l'information)

Procédure de diffusion des logiciels

(2009) TGD : Guide laboratoire pour recenser ses développements logiciels

(2010) TGD : Diffuser un logiciel de laboratoire : recommandations juridiques et administratives

(2014) TGD : Free software, Open source software, licenses. A short presentation including a...

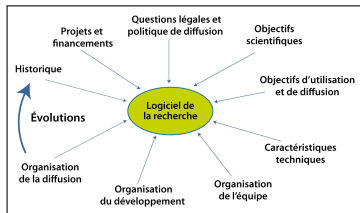
S'adapte à chaque situation, valable pour les données (de la recherche).

- Choisir un nom, éviter les noms déjà utilisés, les marques.
- (*) Établir la liste des auteurs (avec % de participation), leurs affiliations.
- (*) Établir la liste des fonctionnalités principales.
- (*) Établir la liste des briques logicielles ou les données utilisées, avec licences.
- Choisir une licence, avec l'accord des auteurs et propriétaires des droits.
Si possible : un accord signé. Attention à la compatibilité et héritage des licences.
- Choisir un site web, forge, dépôt pour la distribution. Indiquer les licences et les conditions d'utilisation, copie..., comment citer l'œuvre. PIDs...
- Créer et indiquer une adresse courriel de contact.
- (*) La traçabilité est importante, archiver en .tar.gz régulièrement.
- Informer la direction des laboratoires et les tutelles (si pas fait au point licence).
- Diffuser le logiciel et/ou les données.
- Informer la communauté cible, considérer les data ou les software papers.

(*) À revoir à chaque nouvelle version du logiciel.

Plans de gestion de logiciels : modèle PRESOFT

(2018) TGD, G. Romier : Research Software management Plan Template V3.2, Projet PRESOFT



Le modèle est précédé d'une présentation qui indique comment l'utiliser.

0. Titre avec le nom du logiciel, avec :
dates, auteur(s), affiliation(s)

1. Métadonnées

2. Contexte du logiciel →

3. Caractéristiques du logiciel →

4. Organisation de l'équipe

5. Organisation du développement

6. Organisation de la diffusion

7. Gestion du plan de gestion

Références

2.1 Historique

2.2 Projet(s) liés au logiciel

2.3 Questions légales et politique de diffusion

3.1 Objectifs scientifiques

3.2 Objectifs d'utilisation et de diffusion

3.3 Caractéristiques techniques

Études : questions juridiques et de politique scientifique

(2009) TGD : Licence & copyright pour les développements de logiciels libres de laboratoires...

(2009) TGD : Guide laboratoire pour recenser ses développements logiciels

(2011, 2015) TGD : Article vs. Logiciel : questions juridiques et de politique scientifique...

Pour comprendre et expliquer les questions juridiques et de politique scientifique sur les logiciels de la recherche par comparaison avec la production principale des articles de la recherche.

Aspects légaux		
	Article	Logiciel
Droit auteur	droits moraux, droits patrimoniaux	droits moraux réduits , droits pat. dévolus à l'employeur
Œuvre	article	code source, code objet, doc., ...
Auteurs	signataires, même %	notion complexe, pb. légal , établir % de participation
Propriétaires	auteurs, même % cession des droits	tutelles en général, mais dépend du régime salarié , des contrats , ...
Dates	soumission, publication	matériel de conception, versions
Évolution	œuvre indépendante	œuvre indépendante ? il faut revoir auteurs, dates, lic., ...
Travaux préc.	références, citations	briques : compatibilité , héritage lic.
Diffusion	éditeur, web	web, forges, besoin de licence
Droits	lire, citer, ne pas copier	lire, ne pas utiliser , ..., besoin lic.
Licences	droits et obligations, CC (web)	droits et obligations, livres, propriétaires

Aspects relatifs à la politique scientifique		
	Article	Logiciel
Définition (L, T)	ok	à définir
Signature (C, T)	ok, déf. par tutelles	à définir (copyright) associer les laboratoires
Références (L, T)	HAL	PLUME
Liste des œuvres (L, T)	document à jour	document inconnu , PLUME peut être utile
Libre accès (C, L, T, CSI)	politique (+/-) ok, dépôt ok (HAL)	politique (lic.) à définir , dépôt à établir
Validation (C, L, T, CSI)	procédure referee , reproductibilité	à définir , validé au sens PLUME
Qualité/évaluation (C, L, T, CSI)	nb. citations	articles associés, attirer utilisateurs, contrats
Motivation (C, L, T, CSI)	recherche, article	recherche, pas le logiciel
Objet (C, L, T, CSI)	scientifique	3D : scientifique, potentiel de transf. de tech., obj. industriel

(2011) <https://projet-plume.org/ressource/article-vs-logiciel>

(2015) <https://www.societe-informatique-de-france.fr/wp-content/uploads/2015/04/1024-5-gomez-diaz.pdf>

Toutes les références données à la fin traitent des questions liées aux aspects juridiques et de politique scientifique.

Sur la Science Ouverte (1/2)

(2020-21) TGD, T. Recio : Towards an Open Science definition as a political and legal framework: on the sharing...

Trois piliers pour une définition de "Science Ouverte" :

- 🏠 Le libre accès aux publications : déclaration de Budapest (BOAI, 2002)

<https://www.budapestopenaccessinitiative.org/>

Par "accès libre" à cette littérature, nous entendons sa mise à disposition gratuite sur l'Internet public, permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces articles...

- 🏠 Selon la Free Software Fondation (FSF, 1985), fondée par R. M. Stallman, un logiciel est libre <http://www.gnu.org/philosophy/free-sw.fr.html>

si ces quatre libertés sont garanties : d'exécuter, d'étudier et de modifier...

Important : accès au code source, rôle de la licence (conditions du partage).

- 🏠 Mouvement de Open Data : CODATA (Committee on Data for Sciences and Technology) constitué en 1966 par l'International Council of Scientific Unions (ICSU) : http://www.codata.info/resources/newsletters/Newsletter_7.pdf

CODATA est un Comité au niveau scientifique international le plus élevé [...] à cause de l'importance qui s'attache à l'évaluation des données...

Vision + récente : Research Data Alliance (RDA) <https://www.rd-alliance.org/about-rda>

En cours : Recommandation UNESCO sur une Science Ouverte

<https://en.unesco.org/science-sustainable-future/open-science/recommendation>

Projet de recommandation : https://unesdoc.unesco.org/ark:/48223/pf0000378381_fre

Sur la Science Ouverte (2/2)

(2020-21) TGD, T. Recio : Towards an Open Science definition as a political and legal framework: on the sharing...

Définition proposée de "Science Ouverte" :

[EN] the political and legal framework where research outputs are shared and disseminated in order to be rendered visible, accessible and reusable.

[FR] le cadre politique et juridique dans lequel les productions scientifiques sont partagées et diffusées afin d'être rendues visibles, accessibles et réutilisables.

L'article étudie quelques contextes politiques et légales, ainsi que les outils pour faciliter l'adoption de ces nouvelles pratiques :

- Europe : Programme de financement H2020 (2012) *"Towards better access to scientific information : Boosting the benefits of public investments in research"* [PDF](#)
...so that the 'fifth freedom' of the EU – the free circulation of knowledge – can become a reality.
- France : Loi pour une République numérique (7 octobre 2016), et Le Plan national pour la science ouverte (4 juillet 2018)
- Et trois pierres angulaires, les vecteurs du changement :
 - ▶ évolutions des politiques des Universités et des Établissements de recherche,
 - ▶ le développement des services et infrastructures pour la Science Ouverte,
 - ▶ l'évolution des pratiques d'évaluation de la recherche.

Le protocole **CDUR** pour l'évaluation des logiciels de la recherche fait partie de la 3^{ème} pierre.

Comment donner de la valeur à cette production ?

En augmentant la réputation scientifique des producteurs des **RS**, voir :

(2019) TGD, T. Recio : On the evaluation of research software: the CDUR procedure

(C) Citation mesure si RS bien identifié : référence, metadata, bien citer autres...
point légal : auteurs, affiliations, participation

(D) Dissemination bonnes pratiques de diffusion selon contexte de l'évaluation,
[29] (2014) TGD : Free software, Open source software, licenses...
point Science Ouverte, légal : licences

(U) Use aspects "logiciel" **du RS** : résultats corrects, facilite la réutilisation ; et aussi bonnes pratiques logiciel : exemples, docs., test, facilite install., voir le code, lancer le RS, qualité, dépôt APP...
point reproductibilité, validation des résultats obtenus

(R) Research aspects "recherche" : qualité du travail scientifique, algorithmes et structures de données, publications, collaborations et projets...
point impact de la recherche

Flexibilité d'application : chaque décideur ou comité d'évaluation **établit son propre protocole CDUR** adapté au contexte et aux objectifs de l'évaluation.

Conclusions

Sésame, ouvre-toi ! Infrastructures et services pour les logiciels

→ travail en cours au groupe de travail de l'Univ. Gustave Eiffel

Définition Ce que je fais est un logiciel LIGM ?

→ définitions proposées : logiciels d'un laboratoire, de la recherche

Référence Pas de nom, pas d'auteurs, pas de date, pas de version...

→ référence, procédures de diffusion, plans de gestion des logiciels

Diffusion Quelques uns sur pages web perso, sur des forges, avec page web...

→ procédures de diffusion, plans de gestion des logiciels

FLOSS Un réel intérêt pour les logiciels libres, mais souvent pas de licence

→ besoin de continuer la formation, le support : nouveaux membres...

Questions légales Qui décide de la licence ?

→ besoin de continuer la formation, le support : nouveaux membres...

Questions de politique Définition, licences, diffusion, décisions...

(2009) TGD : Guide laboratoire pour recenser ses développements logiciels

→ travail en cours au groupe de travail de l'Univ. Gustave Eiffel

Valeur Quelle est la valeur de cette production scientifique ?

→ proposition des protocoles d'évaluation CDUR

Questions de nature diverse, similaires pour tout domaine scientifique :

→ solutions adaptables proposées.

Références

- 2007 TGD : Autour de la valorisation de logiciels développés dans un laboratoire de recherche.
- 2009 TGD : Licence & copyright pour les développements de logiciels libres de laboratoires de recherche, PLUME, <https://projet-plume.org/fr/ressource/faq-licence-copyright>
- 2009 TGD : Guide laboratoire pour recenser ses développements logiciels, PLUME, <https://projet-plume.org/ressource/guide-laboratoire-recensement-developpements-logiciels>
- 2010 TGD : Diffuser un logiciel de laboratoire : recommandations juridiques et administratives, PLUME, <http://www.projet-plume.org/fr/ressource/diffuser-logiciel-recomm-juridiques-admin>
- 2014 TGD : Free software, Open source software, licenses. A short presentation including a procedure for research software [...] dissemination, Zenodo, <https://zenodo.org/record/11709>
- 2015 TGD : Article vs. Logiciel : questions juridiques et de politique scientifique dans la production de logiciels, Société Informatique de France, <https://zenodo.org/record/18993> (2011) V1, PLUME, <http://www.projet-plume.org/fr/ressource/article-vs-logiciel>
- 2018 TGD, G. Romier : Research Software management Plan Template V3.2, Projet PRESOFT, Zenodo, <https://zenodo.org/record/1405614>. Voir aussi sur [DMP OPIDoR](#).
- 2019 TGD : Le Projet PLUME et le paysage actuel des logiciels de la recherche dans la science ouverte, Zenodo, <https://zenodo.org/record/2591474>
- 2019 TGD, T. Recio : On the evaluation of research software: the CDUR procedure, F1000Research, <https://doi.org/10.12688/f1000research.19994.2>
- 2020-1 TGD, T. Recio : Towards an Open Science definition as a political and legal framework: on the sharing and dissemination of research outputs, POLIS N. 19, 2020, <http://uet.edu.al/polis/images/1.pdf>, V3 du 28/02/2021 : <https://zenodo.org/record/4577066>
- 2021 TGD, T. Recio : Open comments on the Task Force SIRS report: Scholarly Infrastructures for Research Software, RIO 7 : e63872, <https://doi.org/10.3897/rio.7.e63872>